Fall 2009

Sparse representation for audio noise removal using zero-zone quantizers

Neha Mittal
New Jersey Institute of Technology

ABSTRACT

SPARSE REPRESENTATION FOR AUDIO NOISE REMOVAL USING ZERO-ZONE QUANTIZERS

by
Neha Mittal

In zero zone quantization, bins around zero are quantized to a zero value. This kind of quantization can be applied on orthogonal transforms to remove the unwanted or redundant signal. Transforms reveal structures and properties of a signal and hence careful application of a zero zone over the transform coefficients leads to noise removal. In this thesis, such quantizers are applied over Discrete Fourier Transform and Karhunen Loeve Transform coefficients separately, and outputs compared. Further, the localization of the zero zones to certain frequencies leads to better performance in terms of noise removal. PEAQ (Perceptual Evaluation of Audio Quality) scores have been used to measure the objective quality of the denoised signal.

SPARSE REPRESENTATION FOR AUDIO NOISE REMOVAL USING ZERO-ZONE QUANTIZERS

by
Neha Mittal

A Thesis
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Master of Science in Electrical Engineering

Department of Electrical and Computer Engineering

January 2010

APPROVAL PAGE

SPARSE REPRESENTATION FOR AUDIO NOISE REMOVAL USING ZERO-ZONE QUANTIZERS

Neha Mittal

12/8/09

Dr. Ali N. Akansu, Thesis Co-Advisor — Date
Professor of Electrical and Computer Engineering, NJIT

12/09/09

Dr. Richard A. Haddad, Committee Member — Date
Professor of Electrical and Computer Engineering, NJIT

12/08/09

Dr. Deepen Sinha, Thesis Co-Advisor — Date
CEO, ATC-Labs, Newark, NJ

BIOGRAPHICAL SKETCH

Author: Neha Mittal

Degree: Master of Science

Undergraduate and Graduate Education:

- Master of Science in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, 2010

- Bachelor of Technology in Electronics and Communication Engineering,
  Indira Gandhi Institute of Technology, Delhi, India, 2007

Major: Electrical Engineering

Dedicated to
My Family and Friends

ACKNOWLEDGMENT

I would like to express my deepest gratitude to my thesis advisors; Dr. Ali Akansu and Dr. Deepen Sinha for the opportunity to work under their guidance and for their constant support and encouragement. Prof. Akansu not only provided valuable resources but also gave me countless ideas to achieve my objective. My appreciation also goes out to Dr. Sinha and his team in ATC Labs who helped me to implement my ideas and see them in life.

I would like to also thank Dr. Richard Haddad, for being in my thesis committee. His great teaching in my coursework really formed the basis for my efforts in this research area.

I would like to thank Handan Agirman and Mustafa Ugur Torun for sharing with me their vast knowledge and experience with the subject.

Last but not the least, I would like to express my appreciation to my family especially my aunt Nupur Mittal for sharing with me her experiences on research and writing a thesis. I feel grateful to my friends, Lav, Aparna and Tanu who were a great support and boosted my confidence throughout my academic stay at NJIT.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(Continued)

CHAPTER 1

INTRODUCTION

1.1 Objective

Signal contamination by noise can be quite a menace in various fields like broadcasting, telecommunications, forensic science etc. The output of noise reduction algorithms follow the statistics of the input signal, calling for a need to design different type of techniques for different noise and signal statistics. Noise removal has called for a lot of research and a simple way to get rid of noise is to threshold the noisy signal to remove signal below a defined threshold. Such thresholds are mostly applied to the signal in their transform domain as the transform reveals the property and the structure of the signal. Orthogonal transforms decorrelate the signal and repack their energy into less number of coefficients, leading to a sparse representation of the signal. This thesis aims at designing and then comparing several sparse representation techniques which result in noise reduction.

1.2 Background Information

The ongoing research has lead to numerous noise removal techniques depending on the area of application. Techniques might apply processing in the time domain or frequency domain or both. Spectral Subtraction [1] is one of the most commonly used time domain based techniques. In spectral subtraction, the short time power spectrum of the noisy signal is computed and the short time power spectrum of the noise is subtracted from it to receive the estimate of clean signal. However, some of the known issues with this algorithm lie in the estimation of the noise spectrum. Weiner filtering and Kalman

filtering are two of the widely used frequency domain techniques for noise removal. Weiner filtering reduces the noise content by comparing the output with an estimate of the desired noiseless signal but these computations appear to be more efficient for short processing blocks. Kalman filtering is theoretically the most efficient recursive filter but the computational complexities come in. Effective noise removal can be achieved from Spectral Subtraction by using a Voice Activity Decoder (VAD) which estimates the properties of noise from the processed samples in past frames.

Such techniques call for complex and sophisticated implementations. Simple and robust theoretical concepts can be used to understand the properties of the actual signal and noise, careful thresholding can lead to perceptual loss of noise from the data. An attempt to get rid of the noise can also lead to parts of the actual signal getting distorted. Hence, the level and quality of noise removal differs depending on application.

1.3 The Scope of Thesis

This work presents several sparse representation techniques that lead to noise removal. Chapter 2 introduces the concept of sparse representation and several signal analysis techniques. Chapter 3 talks about noise removal and tools for objectively and subjectively measuring the quality of the speech. Chapter 4 presents the methods designed for this thesis, results and their implication. Finally, Chapter 5 concludes the thesis while drawing some conclusions.

CHAPTER 2

SPARSE REPRESENTATION AND ANALYSIS TECHNIQUES

The enormity of the amount of data generated and transmitted these days' demands efficient data compression techniques. Data compression encodes information using smaller number of bits than a regular representation requires. Compression is mainly used to get rid of the signal redundancies and low energy component of the signal. The simplest compression algorithm can comprise of getting rid of some samples of data completely by applying thresholds. Careful application of thresholds can be used to the benefit of noise removal. This Chapter introduces a very elementary technique of achieving compression, Sparse Representation. It further introduces some signal analysis techniques including orthogonal transforms and subband decomposition.

2.1 Sparse Representation

Sparse representation uses a linear combination of less number of elementary signals to represent the vast amount of information present in a signal [2]. The collection of elementary signals whose linear combination is used to represent the signal is known as a dictionary and the elements of this dictionary are known as atoms.

Let a signal x $\in R^N$ can be represented as:

$$x = \boldsymbol{\phi}\,\theta \tag{2.1}$$

where $\theta \in R^N$. $\boldsymbol{\phi}$ is a vector each value of which is an element of a dictionary.

Dictionaries can be orthogonal basis such that $\boldsymbol{\phi}$ depicts the kernel and $\theta = \boldsymbol{\phi}^{-1}x$ are the unique transform expansion coefficients.

The representation is exact-sparse if a large number of the coefficients θ are zero. This representation is made sparser by applying a threshold such that it rejects the coefficients with values around zero. Sparse representation finds its application in compression applications where the information to be transmitted gets reduced through thresholding.

Sparse representations have been known to provide extremely high performance for applications such as noise removal, compression, feature extraction, pattern classification and blind source separation. Lossy compression leads to noise removal [6]: noise suppressing effect and thresholding allows for simultaneous noise reduction and compression.

2.1.1 Quantization

Quantization is the process of approximating a continuous range of values by a finite and smaller set of discrete values. It reduces the number of bits needed in order to store an integer value by reducing the precision of the integer. In quantization, several quantization levels are obtained which estimates the number of bits required to transmit the data.

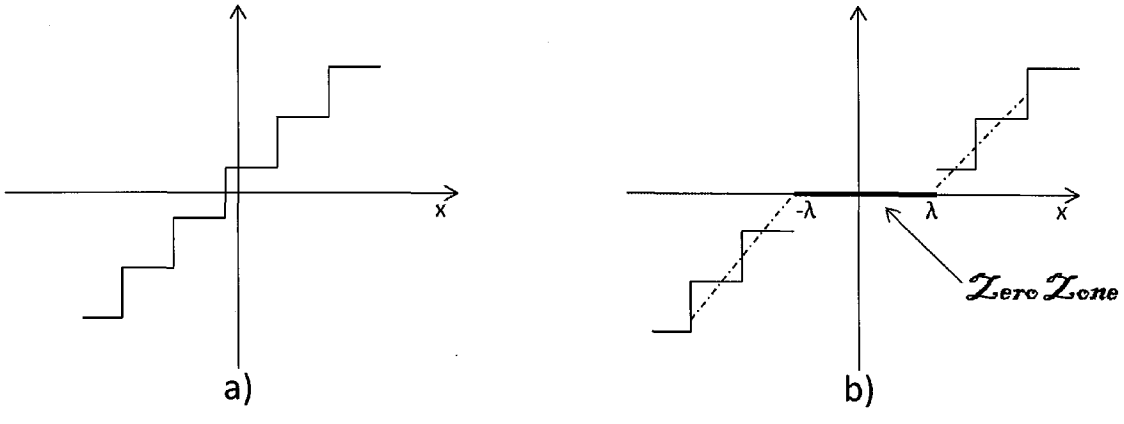


Figure 2.1 a) Uniform Quantization b) Zero Zone Quantization.

Quantization can be classified into Scalar or Vector Quantization depending on whether the quantization is performed on individual coefficients or a group of coefficients [3]. In scalar quantization, quantization bins are formed out of the input data X. As explained in [4], suppose the input data X ranges within the interval [a,b], this interval can be broken down into say 'm' bins. Each bin would carry varying number of data points. If the interval [a, b] is divided into bins of equal size, it is said to be uniform quantization (Figure 2.1(a)). A threshold is then decided to bring some of the bins to zero value and the rest are transmitted as it is (Figure 2.1 (b)).

The human auditory system is less sensitive to the unvoiced speech segment compared to the voiced speech [5]. For the voiced speech, the power spectrum displays a more harmonic structure between frequency intervals of 75 to 3.5 kHz, which is called the voiced band; this interval varies from speaker to speaker. When the harmonic structure doesn't exist in the power spectrum, the speech is called unvoiced. In time-domain such segments play a noise – like structure. However, many regions of the natural speech display a combination of voiced and unvoiced speech, i.e. a combination of harmonic spectrum and noise spectrum. A lower bit rate is required for the unvoiced signal because the human auditory system can't distinguish different noise-like signals. Hence quantization or reduction in the accuracy of the higher spatial frequency values does not dramatically affect the speech quality, since lower frequency signal is unchanged.

Quantization where parts of the higher frequency coefficients are not transmitted is considered zonal sampling. (In most scenarios, variations around zero are due to the

effect of noise on the data.) In some quantization process, the threshold level is increased around zero and is called a dead-zone or a zero-zone quantization. Since, variations around zero are due to the effect of noise on the data, such a quantization results in data compression along with noise removal. Quantization with a dead-zone around zero actually eliminates the noise around zero and improves the signal quality.

2.1.2 Zero Zone Quantization

In zero zone quantization, the transform coefficients of the signal are mapped to a fixed number of bins and thresholding is applied. Thresholding can be obtained by quantizing certain bins around zero to a zero value. This quantization would in effect reduce the precision of the coefficients and in effect map insignificant coefficient values to zero whilst reducing the number of significant, non-zero coefficients that are being used to represent information. According to information theory, the required average bit allocation per data bin is given by

$$E = -\sum_{i} P_i \ log_2(P_i)\,, \tag{2.2}$$

where the labels P_i indicate the probability of the distribution at the bin location i. Hence the output of this quantizer gives a sparse representation of the input signal, mainly containing zeros.

Zero zone quantization also leads to reduction in noise levels as large variations around zero are mostly due to addition of noise to the signal. By applying a zero zone threshold, noisy transform coefficients can be removed which would in effect improve the Signal - to – Noise Ratio (SNR).

It is known that human ear uses both temporal and spectral components to perceive the audio signal. An audio signal requires to be analyzed in terms of the frequency- domain characteristics as well as the time- domain characteristics. However, the Heisenburg's uncertainty states that the time and frequency component can't be both known at the same time with total certainty.

This thesis compares the effect of applying the zero zone quantization on different transform bases. Dead zone quantization is applied on two different transform coefficients. Due to the properties of block transforms, the signal is affected in different ways.

2.1.3 Choosing Threshold

The choice of the threshold depends on the spread of the signal onto which the threshold is being applied. The common idea here is that the less significant transform coefficients are forced to zero. So, choosing a threshold around zero value makes more sense.

Figure 2.2 shows the probability density function (pdf) of a signal. Threshold λ can be defined as $\lambda = \mu + \alpha \cdot \sigma$ where μ is the mean of the pdf and σ is the standard deviation. This threshold is used to decide a level, below which all the coefficients are set to zero.

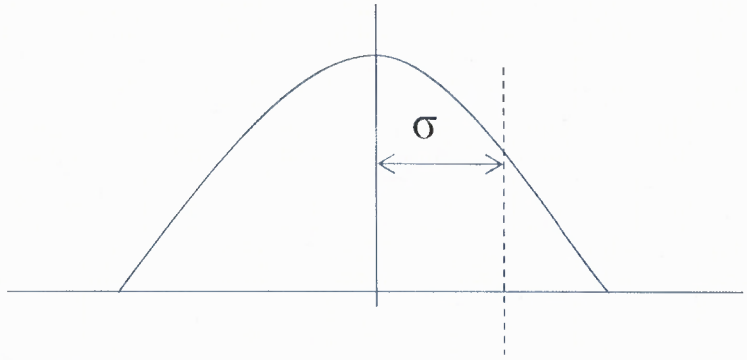


Figure 2.2 Choosing threshold from the signal pdf.

2.2 Signal Analysis Techniques

2.2.1 Discrete Fourier Transform (DFT)

Applying DFT on a signal gives the Fourier transform of the sampled signal s(n) with a finite number of samples (say N). When K is the size of the DFT, it is defined as:

$$S(k) = \sum_{n=0}^{N-1} s(n)e^{-j2\pi nk/N} \quad where\ k \in \{0, \ldots K-1\} \tag{2.3}$$

DFT allows the computation of spectra from discrete-time data. Consider a DFT of size 512, Figure 2.3 shows the magnitude plot of DFT of various signals with different noise levels. It can be observed that the amplitude between index 100 and 400 is increasing significantly with increasing noise energy, indicating that most of this is noise. Applying a threshold in this region would bring these amplitude values down to zero.

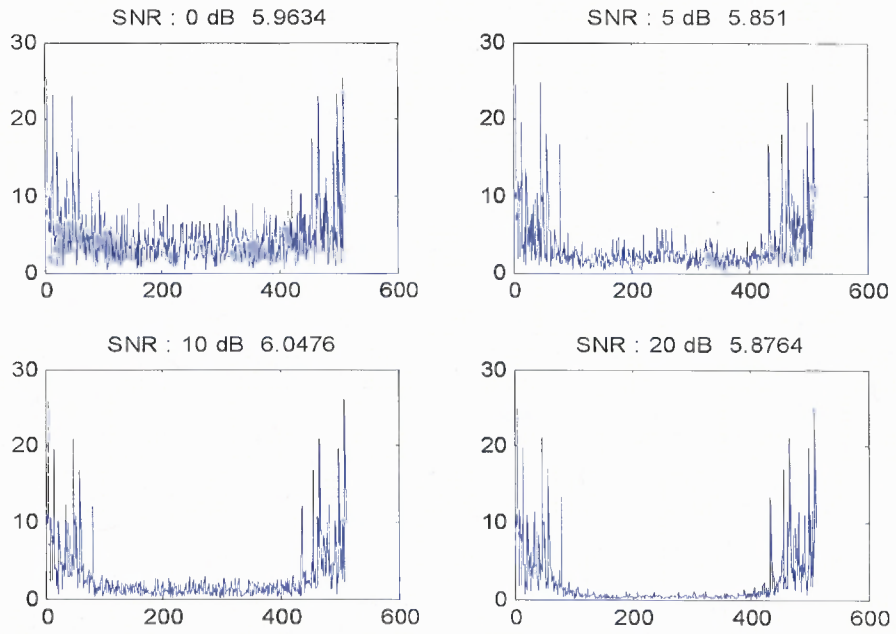


Figure 2.3 Frequency magnitude plots of signals having different SNR.

The architecture of Zero Zone Quantization been applied to the DFT coefficients is shown in Figure 2.4 below. Figure 2.5 shows the effect of applying zero zone on the histogram plot of the DFT coefficients of the signal.

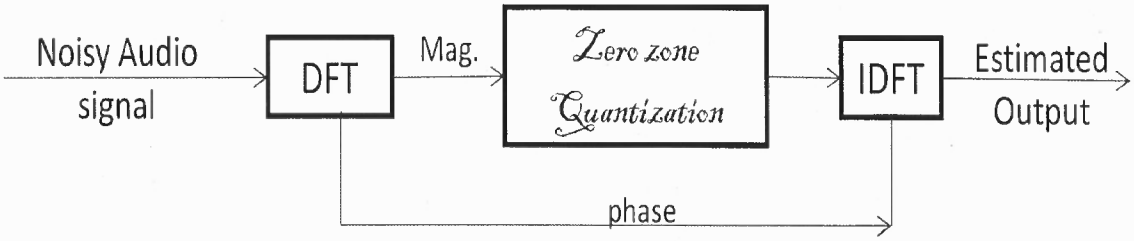


Figure 2.4 Architecture of Zero Zone Quantization on DFT Coefficients.

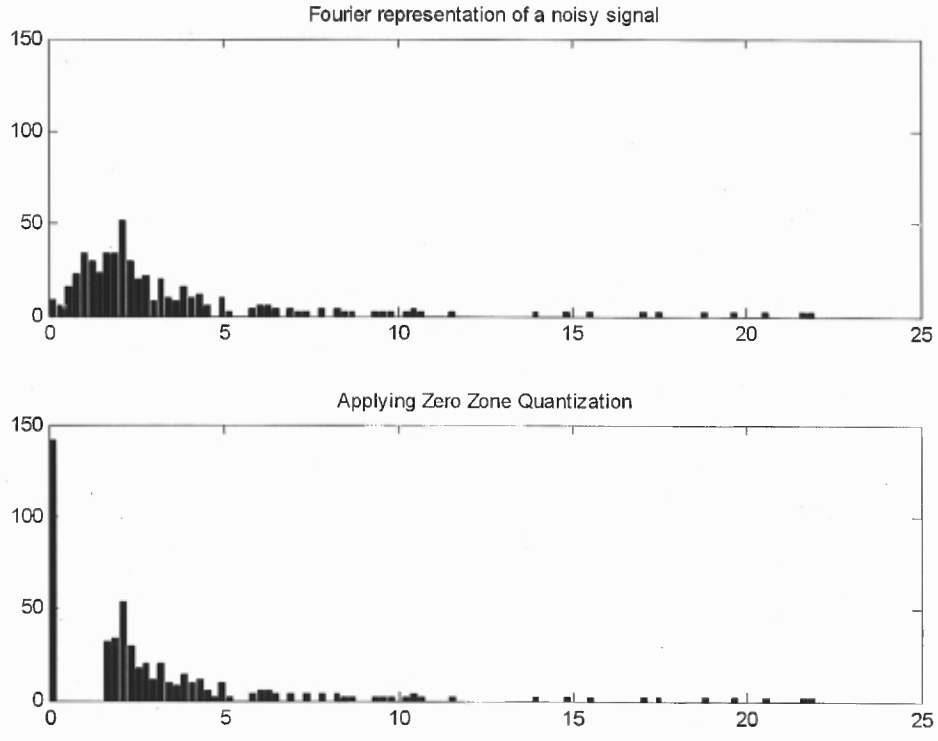


Figure 2.5 Application of Zero zone quantization on DFT Coefficients.

The architecture in Figure 2.4 can be used to plot the rate distortion curves as shown in Figure 2.6. These curve shows the distortion with respect to the distorted signal. The plots depict how the transmission rate decreases with the increasing thresholds. The rate is defined as the average entropy of signal (Equation 2.2) over the frames and the distortion is the mean square error between the input and the output signal.

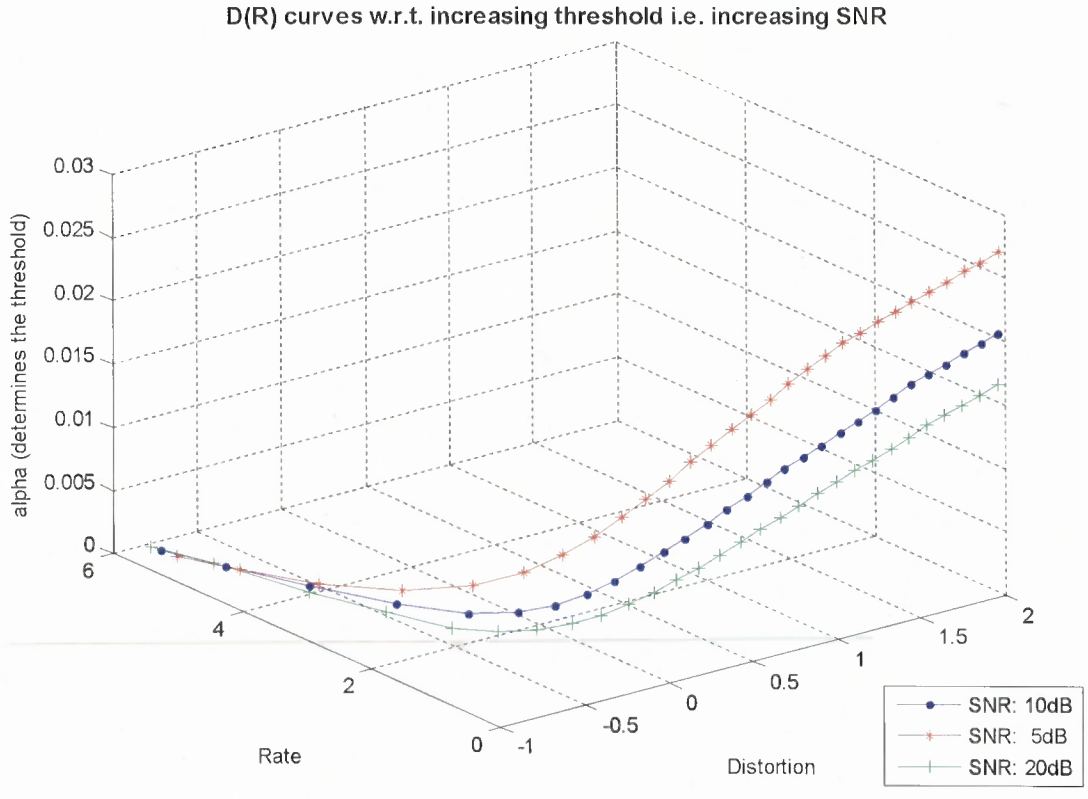


Figure 2.6 D(R) curve for DFT.

2.2.2 Karhunen Loeve Transform (KLT)

KLT diagonalizes the covariance or the correlation matrix of a discrete random sequence and is considered an optimal transform [6]-[9]. KLT depends on the statistics of the signal and hence its computational complexity is very high. Such a transform is also called principal component.

$\boldsymbol{R}$ denotes the $(N \times N)$ correlation matrix of a random complex sequence

$\boldsymbol{x} = (x_1, x_2, x_3, \cdots, x_N)^T$ given by

$$\boldsymbol{R} = E[\boldsymbol{x}\boldsymbol{x}^{\mathrm{H}}] = E\begin{bmatrix} x_1x_1^* & x_1x_2^* & x_1x_3^* & \cdots & \cdots & \cdots & x_1x_N^* \\ x_2x_1^* & x_2x_2^* & x_2x_3^* & \cdots & \cdots & \cdots & x_2x_N^* \\ x_3x_1^* & x_3x_2^* & x_3x_3^* & \cdots & \cdots & \cdots & x_3x_N^* \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_Nx_1^* & x_Nx_2^* & x_Nx_3^* & \cdots & \cdots & \cdots & x_Nx_{N*} \end{bmatrix} \tag{2.4}$$

where E is the expectation operator and $E\lfloor x_j x_j^* \rfloor$ is the autocorrelation of $\lfloor x_j \rfloor$ and $E\lfloor x_j x_k^* \rfloor$ is the crosscorrelation between $\lfloor x_j \rfloor$ and $[x_k]$, j $\neq$k. Note that $\boldsymbol{R}$ is Hermitian. Let the unitary matrix which diagonalizes $\boldsymbol{R}$ be defined as $\boldsymbol{\Phi}$ such that

$$\boldsymbol{\Phi}^{-1} = \boldsymbol{\Phi}^H, \boldsymbol{\Phi}\boldsymbol{\Phi}^H = \boldsymbol{I}, \tag{2.5}$$

$$\boldsymbol{\Phi}^H \boldsymbol{R}\boldsymbol{\Phi} = \boldsymbol{\Phi}^{-1}\boldsymbol{R}\boldsymbol{\Phi} = \boldsymbol{\Lambda}, \tag{2.6}$$

$$\boldsymbol{R}\boldsymbol{\Phi} = \boldsymbol{\Phi}\boldsymbol{\Lambda} \tag{2.7}$$

Equation 2.7 is the characteristic equation, where λ_i, $i = 1,2,3,...N$ are the eigenvalues of $\boldsymbol{R}$. $\boldsymbol{\Phi}$ is called the KLT matrix and it decorrelates the random sequence $\boldsymbol{x}$. The forward KLT of x is written as:

$$\boldsymbol{\theta} = \boldsymbol{\Phi}^{-1}\mathbf{x} = \boldsymbol{\Phi}^{H}\mathbf{x} \tag{2.8}$$

and the inverse KLT transform is given by

$$\mathbf{x} = \boldsymbol{\Phi}.\boldsymbol{\theta} \tag{2.9}$$

where, $\theta = (\theta_1, \theta_2, \ldots, \theta_N)^T$ represents the random sequence in the transform domain.

$$\boldsymbol{\Phi} = [\emptyset_1 \quad \emptyset_2 \quad \cdots \quad \emptyset_{N-1}] = \begin{bmatrix} \Phi_{11} & \cdots & \Phi_{N1} \\ \Phi_{12} & & \\ \Phi_{13} & \vdots & \vdots \\ \vdots & & \\ \Phi_{1N} & \cdots & \Phi_{NN} \end{bmatrix} \tag{2.10}$$

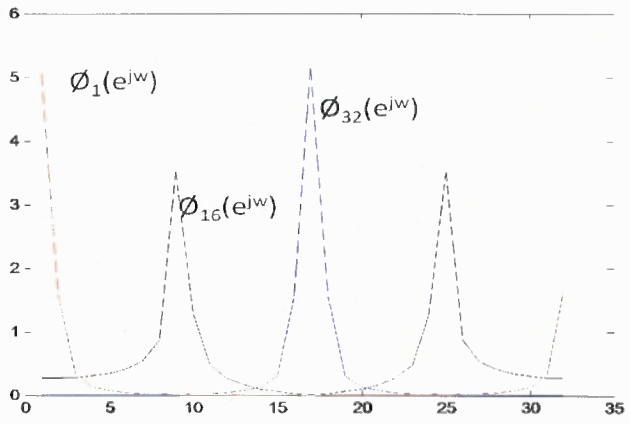


Figure 2.7 Plot of the amplitude of the KLT matrix.

Consider a size 32 KLT, then KLT coefficients, θ can be obtained as in Equation 2.8 in which N=32. Figure 2.8 depicts the magnitude plot of the KLT coefficients. It is clear from this that the higher theta coefficients are low in the energy content.

$$\begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{N-1} \end{bmatrix} = \begin{bmatrix} \Phi_{11} & \cdots & \Phi_{N1} \\ \Phi_{12} & & \\ \Phi_{13} & \ddots & \vdots \\ \vdots & & \\ \Phi_{1N} & \cdots & \Phi_{NN} \end{bmatrix}^{-1} \times \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{bmatrix} \tag{2.11}$$

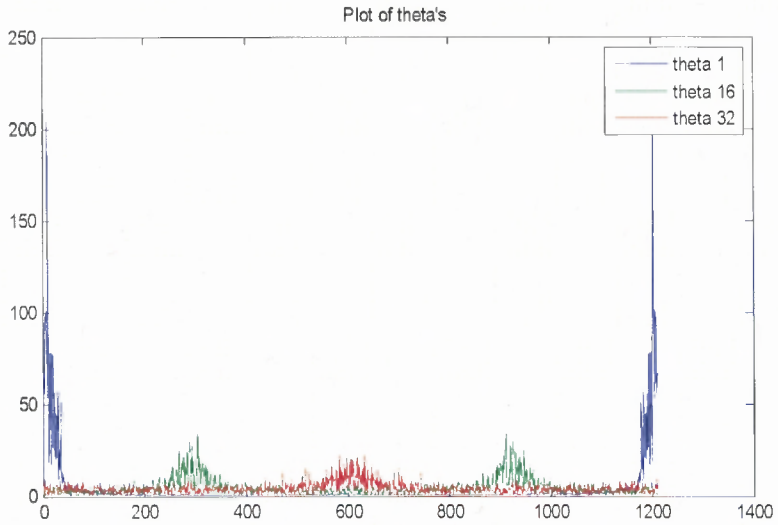


Figure 2.8 Magnitude plot of the KLT coefficients.

Determining the actual correlation matrix ***R*** involves a lot of computation and complexity. An auto-regression (AR) model can be used to model the actual correlation matrix. As the order of the AR model increases, so does its closeness to the actual ***R***. This leads into a better decorrelation of the audio signal.

A simple architecture of the application of zero zone quantization over the KLT transform coefficients is shown in Figure 2.9. The threshold in the case of KLT is defined based on energy criteria, where a percentage of energy is rejected to get rid of low energy coefficients. Figure 2.10 shows the rate distortion curves for this architecture. Here, like the DFT curves, the distortion is defined with respect to the distortion and the rate decreases with the increasing threshold.

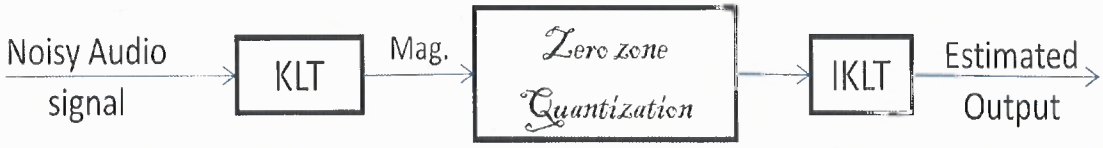


Figure 2.9 Architecture of Zero Zone Quantization on KLT coefficients.

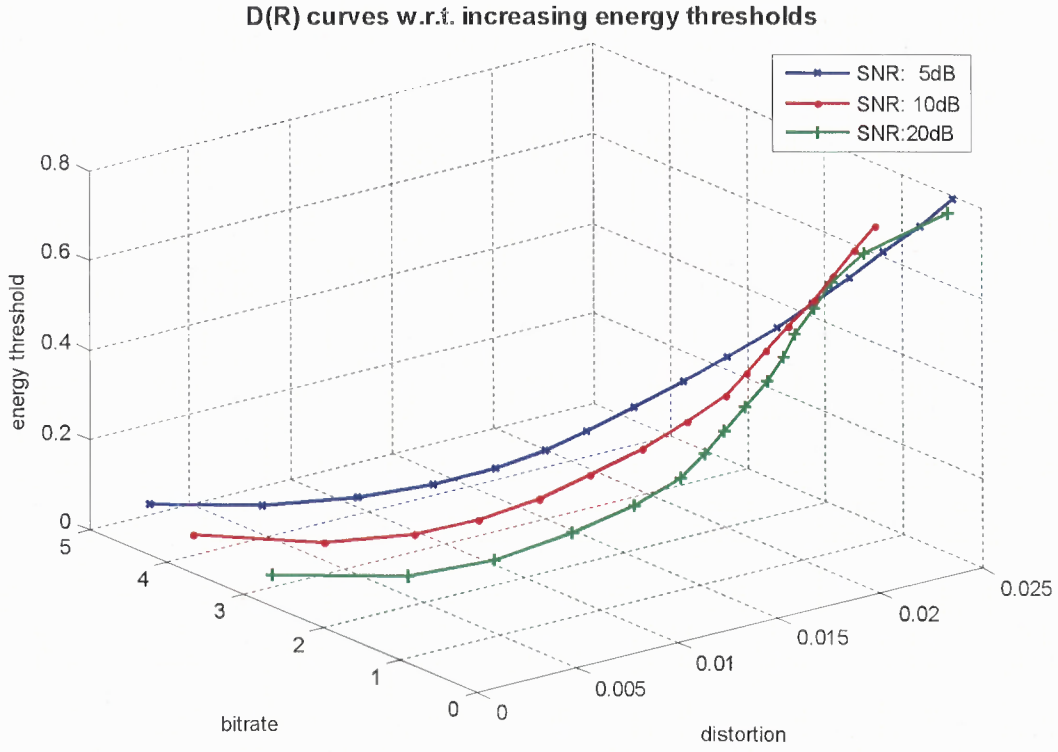


Figure 2.10 D(R) curves for KLT domain thresholding.

2.2.3 Subband Decomposition

The objective of subband decomposition [6], [10] is to divide the signal frequency band into a set of uncorrelated frequency bands by filtering and then passing each of them through a zero-zone quantizer.

2.2.3.1 Analysis Filter Bank. The simplest way to decompose a signal is into one high-frequency component and one low-frequency component. In the filter bank in Figure 2.11, the input signal X(z) is processed simultaneously by a low-pass filter $H_0(z)$ and a high pass filter $H_1(z)$. The available frequency range, from $\omega = 0$ to $\omega = \pi$ (which is half the sampling frequency), is usually partitioned into two halves, as shown in Figure 2.12. The filtered signals thus have a bandwidth of approximately $\pi/2$, and the sampling rate can thus be halved. Small amount of aliasing is accepted.

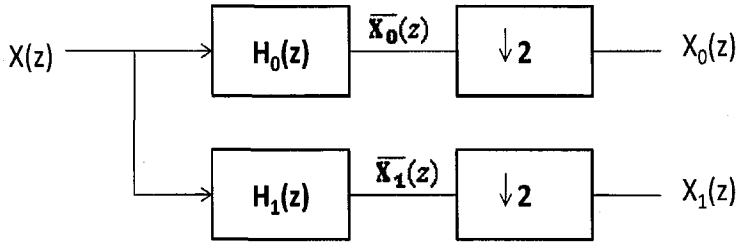


Figure 2.11 Analysis Filter Bank.

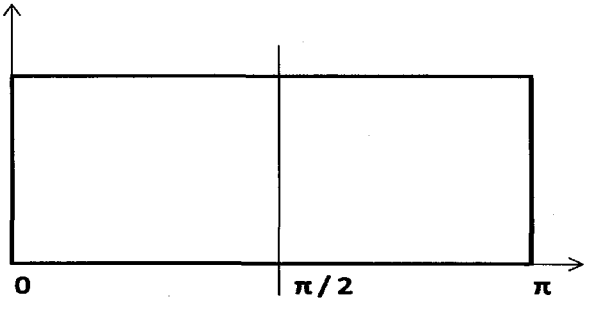


Figure 2.12 Frequency bands with two bands.

The two filtered signals in Figure 2.11 are

$$\overline{X_0}(z) = X(z).H_0(z),$$

$$\overline{X_1}(z) = X(z).H_1(z). \tag{2.12}$$

Downsampling with M=2, yields the subband signals

$$X_0(z) = \frac{1}{2} X\left(z^{1/2}\right) H_0\left(z^{1/2}\right) + \frac{1}{2} X\left(-z^{1/2}\right) H_0\left(-z^{1/2}\right),$$

$$X_1(z) = \frac{1}{2} X\left(z^{1/2}\right) H_1\left(z^{1/2}\right) + \frac{1}{2} X\left(-z^{1/2}\right) H_1\left(-z^{1/2}\right) \tag{2.13}$$

These equations can be written in matrix form as:

$$\begin{bmatrix} X_0(z) \\ X_1(z) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} H_0\left(z^{1/2}\right) & H_0\left(-z^{1/2}\right) \\ H_1\left(z^{1/2}\right) & H_1\left(-z^{1/2}\right) \end{bmatrix} \cdot \begin{bmatrix} X\left(z^{1/2}\right) \\ X\left(-z^{1/2}\right) \end{bmatrix} \tag{2.14}$$

The spectral components in Equations 2.13, which depend on $X\left(z^{1/2}\right)$, lie in the baseband. The spectral components that are a function of $X\left(-z^{1/2}\right)$ are periodic repetations of these. Since the filtered signals are not properly band-limited to π, alias signals appear in the baseband.

2.2.3.2 Synthesis Filter Bank

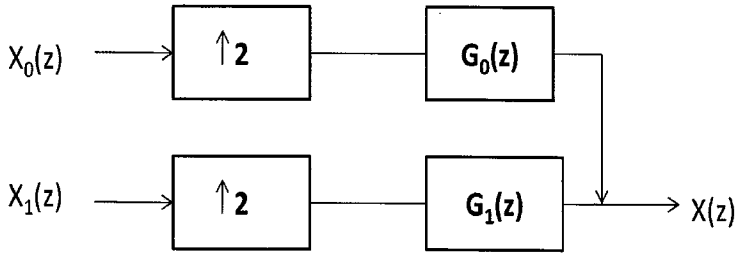


Figure 2.13 Synthesis Filter bank.

Figure 2.13 shows a two-channel synthesis filter bank with a low-pass filter $G_0(z)$ and high-pass filter $G_1(z)$. This is the dual of the analysis filter bank in Figure 2.11. The fundamental characteristics of the two filters are the same as the analysis filter bank. After upsampling the subsignals $X_0(z)$ and $X_1(z)$, the low-pass filter $G_0(z)$ eliminates substantial parts of the spectrum of the low-pass signal $X_0(z)$ in the range $\pi/2 < \Omega < \pi$. The high-pass filter $G_1(z)$, meanwhile, eliminates most of the spectra of the high pass signal $X_1(z)$ that are in the range $0 < \Omega < \pi/2$. Since the frequency responses of the two signals overlap, the signal spectra are not totally eliminated.

The output signal X(z) of the synthesis filter bank can be expressed as

$$X(z) = G_0(z).X_0(z^2) + G_1(z).X_1(z^2), \tag{2.15}$$

This relationship can be expressed in matrix form as:

$$X(z) = [G_0(z) \quad G_1(z)] . \begin{bmatrix} X_0(z^2) \\ X_1(z^2) \end{bmatrix} \tag{2.16}$$

2.2.3.3 Two-channel SBC filter bank. An analysis filter bank followed by a synthesis filter bank, together form a subband coding (SBC) filter bank. Figure 2.14 shows a two-channel SBC filter bank. An analysis filter bank with filters $H_0(z)$ and $H_1(z)$ decomposes the input signal X(z) into the subband signals $X_0(z)$ and $X_1(z)$, this is followed by a synthesis filter bank with filters $G_0(z)$ and $G_1(z)$, which reconstructs the output signal $\widehat{X}(z)$ from the subband signals.

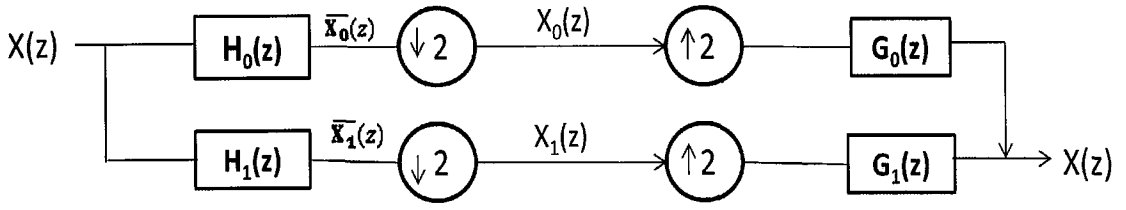


Figure 2.14 Two channel SBC filter bank.

In critically sampled filter banks, the number of samples per unit time of the whole set of subband signals is equal to the number of input samples per unit time. However, the power of the subband signals is generally lower that the original signal. Therefore, for the given quantization error, there is a coding gain if the subband signals is coded instead of the original signal. The coded signal can be used for storage or transmission. After the decoding, the original signal is to be reconstructed. These results in a fundamental requirement for SBC filter banks: analysis and synthesis filter banks should be designed to produce an output signal $\hat{X}(z)$ that is as close as possible, or even exactly the same as the original signal X(z). In addition, it is important to have good selectivity in frequency, so that the sum of the power in the subbands is not much more than the power of the original signal.

Substituting Equation 2.14 into Equation 2.16 gives the relationship between the output signal $\hat{X}(z)$and the input X(z):

$$\hat{X}(z) = \begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix} \cdot \frac{1}{2} \begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} \cdot \begin{bmatrix} X(z) \\ X(-z) \end{bmatrix} \tag{2.17}$$

$$\hat{X}(z) = \frac{1}{2}[G_0(z)H_0(z) + G_1(z)H_1(z)].X(z) + \frac{1}{2}[G_0(z)H_0(-z) + G_1(z)H_1(-z)].X(-z)$$

$$= F_0(z)X(z) + F_1(z)X(-z) \quad (2.18)$$

For perfect reconstruction, the output signal must be identical to the input signal $\hat{x}[n] = x[n]$ or $\hat{X}(z) = X(z)$. For this to be true, the function $F_1(z)$ should be set equal to zero. The function $F_0(z)$ describes the transfer characteristics of the filter bank. The function $F_1(z)$ denotes the alias component, which are produced by the overlapping of the frequency responses. If $F_1(z)$ equals zero, an alias-free filter bank is obtained and the remaining function $F_0(z)$ denotes the quality of the reconstruction. If this is merely a delay i.e. $F_0(z) = z^{-k}$, the filter bank performs perfect reconstruction. This can be written as:

$$\frac{1}{2}[G_0(z)\ G_1(z)]\begin{bmatrix} H_0(z) & H_0(-z) \\ H_1(z) & H_1(-z) \end{bmatrix} = [z^{-k}\ 0] \quad (2.19)$$

The analysis and synthesis filter banks are to be chosen such that Equation 2.19 is satisfied as closely as possible, or in the best case exactly. As there are four variables in this equation, there exist many possible solutions to Equation 2.19, one of them being the Quadrature Mirror Filters (QMFs), which will be introduced next.

2.2.3.4 Standard QMF Banks. QMF are two-channel SBC filter banks with power complementary frequency responses. Starting with a suitable low-pass prototype H(z), the following four filters are specified:

$$H_0(z) = H(z) \tag{2.20}$$

$$H_1(z) = H(-z) \tag{2.21}$$

$$G_0(z) = 2H(z) \tag{2.22}$$

$$G_1(z) = -2H(-z) \tag{2.23}$$

Substituting these equations into Equation 2.18 reveals that the condition $F_1(z) = 0$ is satisfied, i.e. the aliasing components cancel each other out. The factor 2 in the synthesis filters exactly cancels out the facto ½ that is introduced by the downsampling. Note that H(-z) is high-pass if H(z) is low-pass. This can be seen by substituting in $z=e^{j\omega}$ and $-z = e^{j(\omega-\pi)}$. Substituting these values in Equations 2.20 and 2.21,

$$H_1(z) = H_0(-z)$$

$$H_1(e^{j\omega}) = H_0(e^{j(\omega-\pi)}) \tag{2.24}$$

Substituting $\omega \rightarrow \frac{\pi}{2} - \omega$, and noting that the magnitude is an even function of ω, leads to:

$$\left|H_1\left(e^{j(\frac{\pi}{2}-\omega)}\right)\right| = \left|H_0\left(e^{j(\frac{\pi}{2}+\omega)}\right)\right| \tag{2.25}$$

The squared amplitude frequency responses are mirror images of each other about the line $\omega = \pi/2$, which leads to the name QMF and is illustrated in Figure 2.15. The frequency responses of the filters H(z) and H(-z) and thus the filters $H_0(z)$, $H_1(z)$, $G_0(z)$ and $G_1(z)$ are power complementary. Hierarchical subband tree structures can be used to construct multiband perfect reconstruction (PR) filter banks and will be illustrated next.

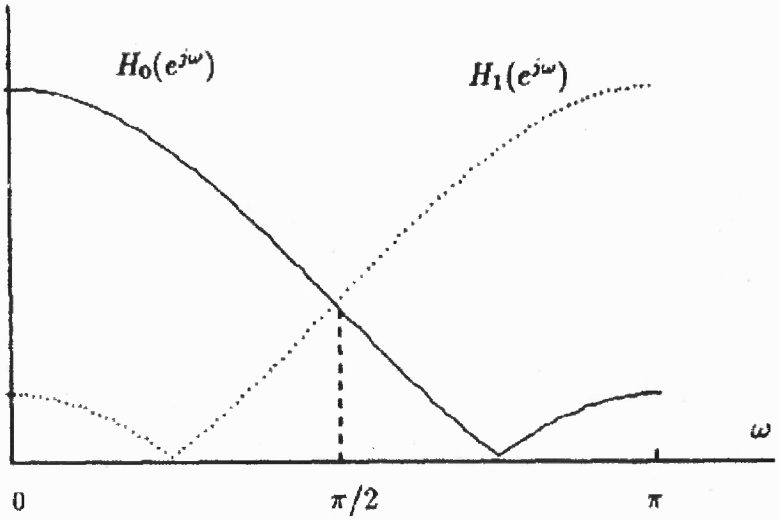


Figure 2.15 Frequency responses of quadrature mirror filters.

2.2.3.5 Regular binary Subband Tree Structure. PR QMF bank can be used for multiresolution spectral analysis. These filter banks divide the input spectrum into two equal subbands, yielding the low (L) and the high (H) bands. This two band PR-QMF split can again be applied to these (L) and (H) half bands to generate the quarter bands: (LL), (LH), (HL), and (HH).

Two levels of this decomposition are shown in Figure 2.16, where the original signal at a data rate less is decomposed into the four subband signals v0(n),..v3(n), each operating at a rate of fs/4. Therefore, the net data rate at the output of the analysis section equals that of the input signal. This conservation of the data rates is called critical sampling.

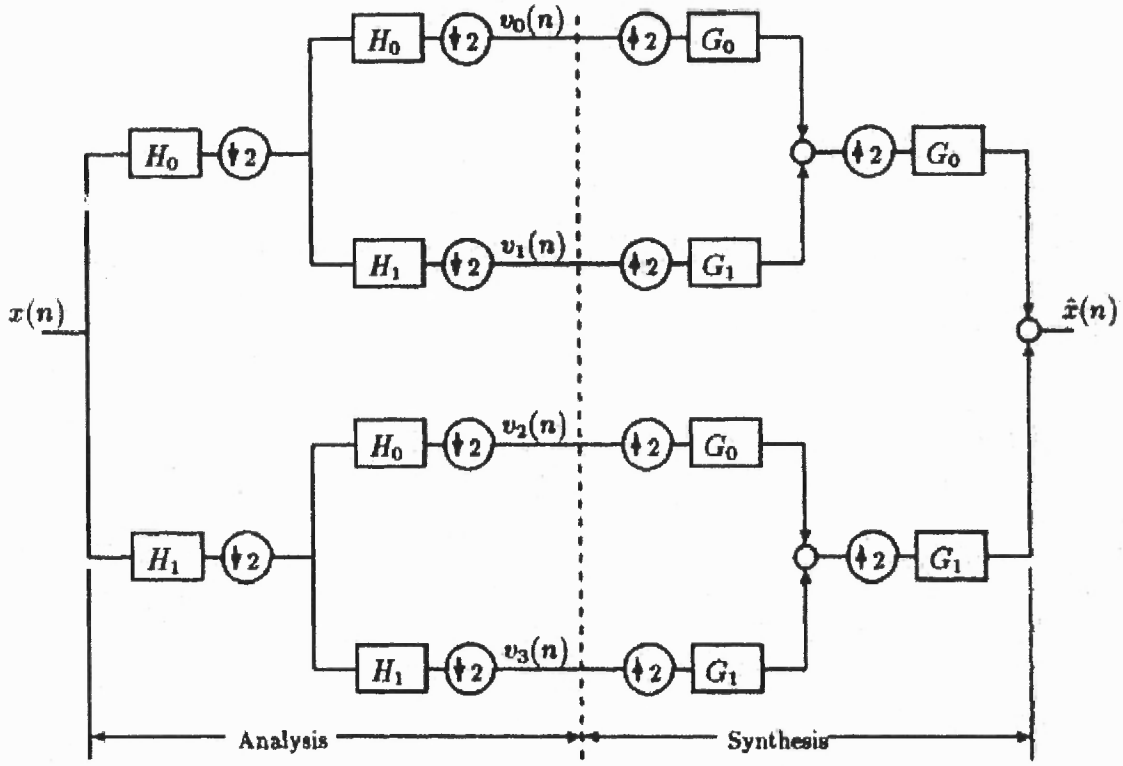


Figure 2.16 Four Band, analysis-synthesis tree structure.

The four band frequency split of the spectrum is shown in Figure 2.17 for ideal band-pass filters.

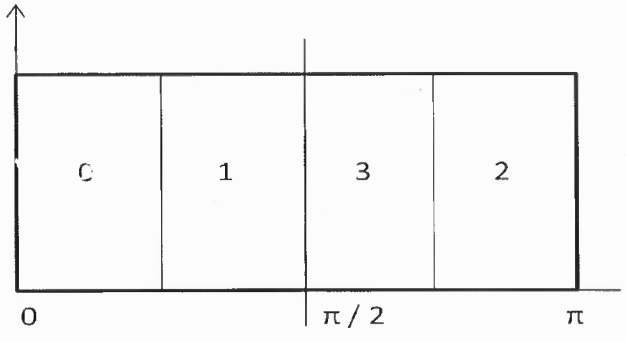


Figure 2.17 Frequency bands corresponding to the four-bands with ideal filters.

2.2.3.6 Dyadic or Octave Band Subband Tree Structure. The Dyadic Analysis Filter Bank decomposes a broadband signal into a collection of subbands with smaller bandwidths and slower sample rates. The dyadic tree is a special irregular tree structure. It splits only the lower half of the spectrum into two equal bands at any level of the tree. Therefore the detail or higher half-band component of the signal at any level of the tree is decomposed no further. The dyadic tree configuration and its corresponding frequency resolution for L=3 are given in Figure 2.18.

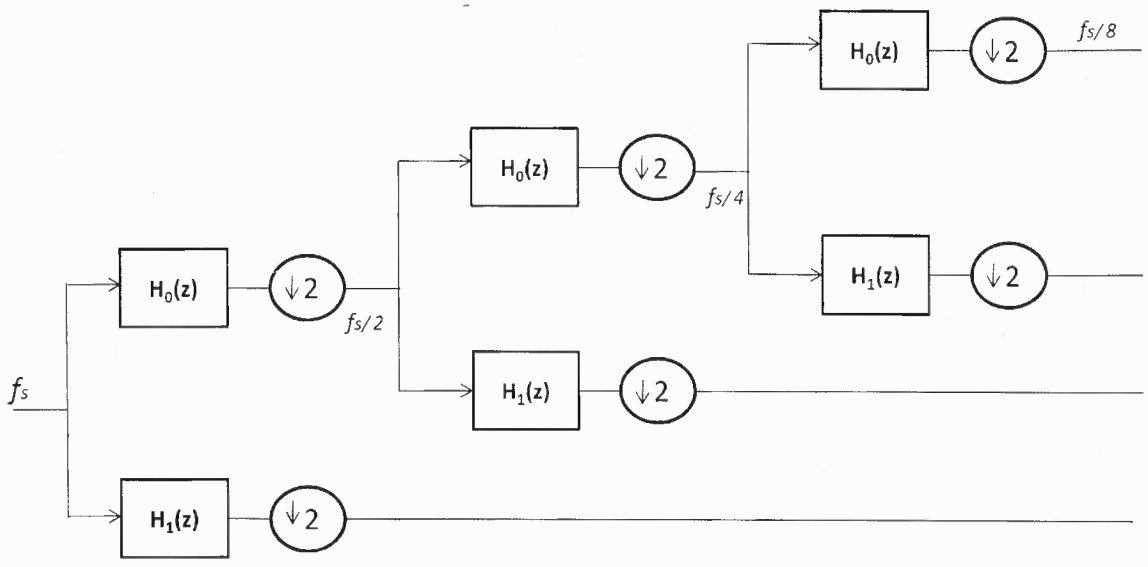


Figure 2.18 Level three Dyadic tree structure.

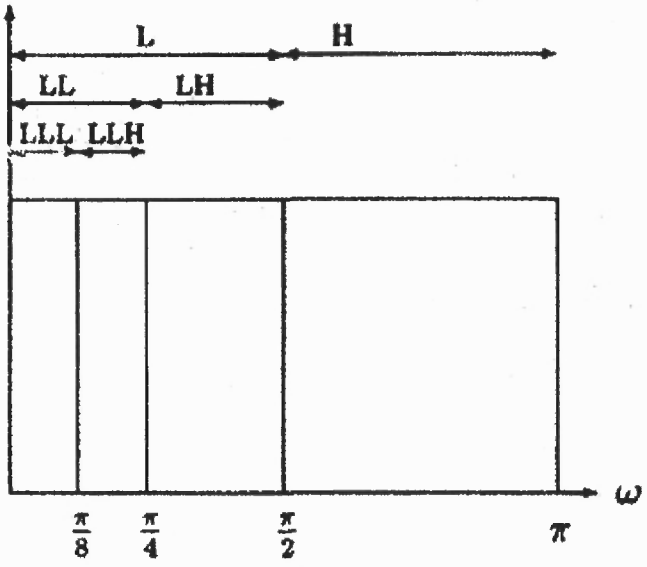


Figure 2.19 Frequency band split corresponding to dyadic tree structure.

A half resolution frequency step is used at each level. Therefore it is also called the octave-based or constant-Q subband tree structure. While the band (L) provides a coarser version of the original signal, band (H) contains the detail information. The approach is repeated L times onto the lower-spectral half component of the higher-level node in the tree.

CHAPTER 3

NOISE REMOVAL

Noise reduction deals with suppressing unwanted noise, thereby enhancing a speech signal in terms of speech quality and intelligibility. In applications like telecommunication and broadcasting, the audio signal often gets contaminated with noise in the form of hiss or buzz due to transmission loss or introduction of background noises. The objective of noise removal is to improve the quality of the signal in order to receive a subjectively comfortable and natural speech at the receiver. A lot of research has been carried in this area which has lead to plenty of techniques which reduce noise based on the properties of either speech or noise.

Usually, it is assumed that the noise is independent of speech signal. Spectral subtraction, which performs subtraction of a noise spectral estimate from a noisy speech spectrum, was suggested in early years, and has still been popular due to its computational efficiency.

While listening to music, one can clearly "hear" the time variation of the sound "frequencies". These localized frequency events are not pure tones but packets of close frequencies. The properties of sounds are revealed by transforms that decompose signals over elementary functions that are well concentrated in time and frequency. Measuring the time variation of "instantaneous" frequencies is an important application that illustrates the limitations imposed by Heisenberg uncertainty.

3.1 Measuring the Quality of the Signal

The quality of the reconstructed speech signal is a vital attribute of a speech coder. It is also a particularly problematic attribute because the evaluation of speech quality is a notoriously difficult problem. As yet, it has not been possible to find an objective criterion that correlates well with speech quality for a variety of speech coders and input signals Furthermore, with decreasing bit rate, the quality of the reconstructed signal of coders becomes more and more dependent on the characteristics of the input signal, making it difficult to anticipate the behavior of the coder in real world applications. Thus, extensive testing with human subjects is required before the suitability of a particular speech coder for a practical application can be judged.

One of the measures that is often used to measure the subjective quality of speech coders is the mean opinion score (MOS). The MOS attempts to combine all aspects of quality in a single number and is perhaps the most commonly used measure for subjective quality testing.

3.1.1 Objective Measures of Quality

Many objective quality measures are used but the signal-to-noise ratio (SNR) in dB is a commonly used measure and is defined as:

$$SNR_{dB} = 10\, log_{10} \frac{\sum s^2}{\sum (s - \hat{s})^2} \tag{3.1}$$

Here, s is the original clean audio signal and $\hat{s}$ is the estimated audio signal. The numerator term is nothing but the energy of the original signal. This approach has a drawback because while removing the noise components from the contaminated signal,

some useful signal components get removed as well. Hence the denominator of the equation 3.1 does not give the actual noise energy.

The quality of the output signal can also be evaluated using spectral distortion (SD) [11]. SD between two signals is calculated as follows:

$$SD = \frac{1}{4I}\sum_{i=1}^{I}\sum_{k=0}^{255} 20(\log_{10} S(k,i) - \log_{10} \hat{S}(k,i)) \tag{3.2}$$

3.1.2 Subjective Measures of Quality

Until recently, the only widely accepted assessment procedures for audio or speech codecs were listening tests, due to the lack of international standards for measuring the perceived audio quality.

The ITU has also recommended a test procedure to assess wide band audio codecs on the basis of subjective tests [11]. Subjective assessments test method focuses on the comparison of the coded/decoded signal to the unprocessed original reference. The relevant recommendation is known as BS.1116, titled "Methods for the Subjective Assessment of small Impairments in Audio Systems including Multichannel Sound Systems" which was issued by the ITU-R2 in 1994 and was updated in 1997.

The test method, which is recommended by BS.1116, is extremely sensitive and allows for the accurate detection of small impairments. The grading scale used is depicted in Table 3.1.

Table 3.1 ITU-R Five-Grade Impairment Scale

Impairment	Grade	SDG
Imperceptible	5.0	0.0
Perceptible,but not annoying	4.0	-1.0
Slightly annoying	3.0	-2.0
Annoying	2.0	-3.0
Very Annoying	1.0	-4.0

The analysis of the results from a subjective listening test is generally based on the Subjective Difference Grade (SDG) which is defined as [12]:

$$\mathbf{SDG = Grade}_{\text{Signal Under Test}} \mathbf{- Grade}_{\text{Reference Signal}} \tag{3.3}$$

Provided that the listener correctly assigns the hidden reference signal, the SDG values will range from 0 to –4, where 0 corresponds to an imperceptible impairment and –4 to an impairment judged as very annoying. The assignment of the SDG scale is shown in the last column in Table 3.1.

3.2 Perceptual Evaluation of Audio Quality (PEAQ)

The idea of substituting the subjective tests by objective, computer based methods gave effect to the development of PEAQ as a standardized algorithm for measuring perceived audio quality in 1994-1998 [13]. It was a joint venture of experts within Task Group 6Q of the International Telecommunication Union (ITU-R). It was originally released as ITU-R Recommendation BS.1387 in 1998 and last updated in 2001. It utilizes software to simulate perceptual properties of the human ear and then, integrate multiple model output

variables (MOV) into a single metric. PEAQ characterizes the perceived audio quality as subjects would do in a listening test according to ITU-R BS.1116. It provides accurate and repeatable estimates of audio quality degradation. It compares the audio signal input to a device under test (DUT) with the corresponding (degraded) audio signal output from that device on a perceptual basis. In perceptual coding it is fundamental to determine the level of noise that can be introduced into a signal before it becomes audible.

OPTICOM who is the leading provider of signal based perceptual measurement technologies and sole licenser for PEAQ offers algorithms for voice, audio and video quality measurements.

3.2.1 OPERA

OPERA™, short for "Objective Perceptual Analyzer", is based on PEAQ to provide an objective evaluation and assure the quality of compressed speech and wide-band audio signals by modeling the human ear. OPERA™ can distinguish between imperceptible and more or less annoying transmission errors as it works quite similar to the human ear.

As a major advantage, OPERA™ employs the same kind of natural stimulus for a measurement as in practical operation: human speech or music program material. As a consequence of the novel approach to measure the perceived audio quality instead of signal characteristics, it is possible to gain an objective quality metrics.

Some of the features of OPERA are at the time:

- ITU-R BS.1387/PEAQ
- Delay measurement
- Real time measurement
- Interfaces to file (*.wav), analog XLR balanced (20 bit) and digital AES/EBU

The common structure of proposed algorithms for perceptual measurement is depicted in Figure 3.1. The process of human perception is modeled by employing a difference-measurement-technique which compares both, a reference signal (i.e. the "input" signal to a codec) and a test signal (i.e. the "output" signal of the codec). First, the algorithms process an ear model for the reference and the test signal, in order to calculate an estimate for the audible signal components.

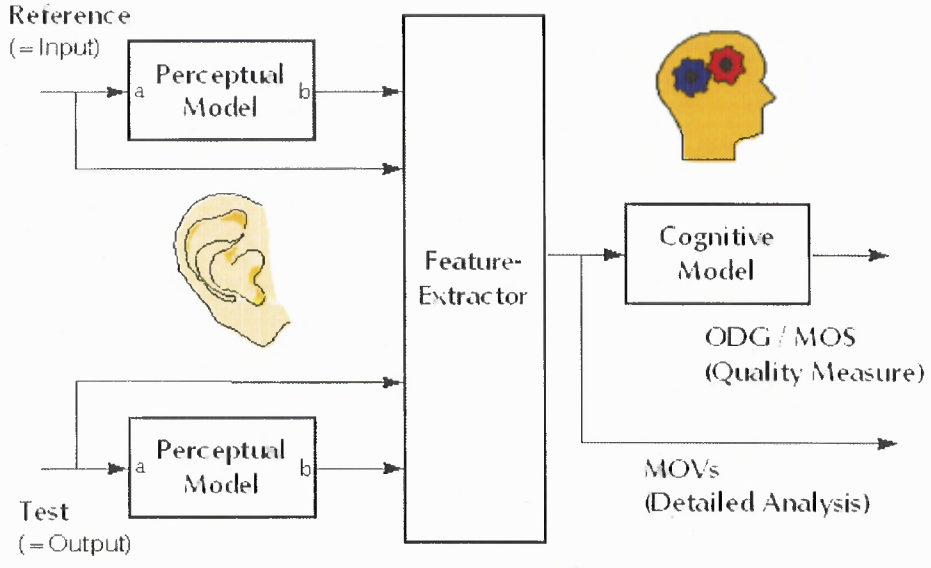


Figure 3.1 The structure of the generic perceptual measurement algorithm.

In a consecutive step, the algorithm models the audible distortion present in the signal under test by comparing the outputs of the ear models. The information obtained by this process results into several values, so called MOVs ("Model Output Variables") and may be useful for a detailed analysis of the signal. The final goal instead is to derive a quality measure, consisting of a single number that indicates the audibility of the

distortions present in the signal under test. To achieve this, some further processing of the MOVs is required which simulates the cognitive part of the human auditory system.

PEAQ in general comprises ear models based on the fast Fourier transform as well as on a filter bank. The output values of the models are based partly on the masked threshold concept and partly on a comparison of internal representations (also known as comparison in the cochlear domain). In addition, it also yields output values based on a comparison of linear spectra, which are not processed by an ear model.

The evaluation of the internal representation is often related to an estimate of the masked threshold. This estimate is based on data found in a number of psychoacoustic experiments. Most of these experiments model certain isolated effects of the human auditory system. One way to design a perceptual measurement algorithm is to generalize these model data and apply them to complex audio signals.

The sample rate of the reference file is frequently already defined by the algorithm that shall be used for the evaluation of the recorded data. PEAQ according to ITU-R BS.1387 requires 48kHz sample rate, although the implementation in OPERA will deliver reliable results at 44.1kHz, too. Most speech quality measures are defined for 8 and 16kHz sample rate only.

OPERA has two versions for PEAQ measurement: The "Basic" version is defined for computational efficiency and realtime performance, while the "Advanced" version yields for highest possible accuracy. The major difference between the Basic and the Advanced version is hidden in the respective ear models and the set of MOVs used. Both versions comprise an artificial neural network for the cognitive modeling. Since these networks are usually critical in terms of reliability, special care was taken not to over-

train the network during the design phase. The "Advanced" version uses some MOVs derived by implementing the ear model of the "Basic" version but in addition to that, the advanced version introduces a second ear model with improved temporal resolution.

3.2.1.1 Model Output Variables. The screen shot in Figure 3.2 shows the Model Output Variables (MOVs) used by the advanced version as they are defined by BS.1387. The results are shown framewise and are averaged since the beginning of the measurement. The interpretation of these MOVs is depicted in the Table 3.2.

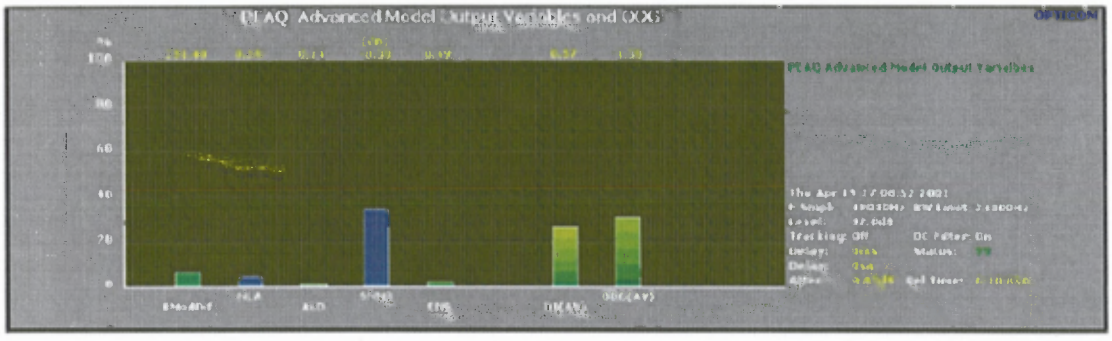


Figure 3.2 PEAQ Advanced Model Output Variables (MOVs) and the ODG.

Table 3.2 MOVs used by the PEAQ "Advanced" version

OPERA Model Output Variable (MOV) Name	**BS. 1387 name**	**Interpretation**
RModDif	$RmsModDiff_A$	Loudness of Distortion
NLA	$RmsNoiseLoudAsym_A$	Changes in modulation (related to roughness)
ALD	$AvgLinDist_A$	Linear distortions (frequency response etc.)
SNMR	Segmental NMR_B	Noise-to-mask ratio
EHS	EHS_B	Harmonic structure of the error

3.2.1.2 Objective Difference Grade (ODG). The last two bars in the diagram shown in Figure 5.24 are the Distortion Index (DI) and the final Objective Difference Grade (ODG). The "AV" in the brackets indicates that this value is a result of the Advanced Version of the PEAQ algorithm. The ODG is the output value from the objective measurement method that corresponds to the SDG (Table 4.1) in the subjective domain. The ODG indicates the measured basic audio quality of the signal under test on a continuous scale from -4 (very annoying impairment) to 0 (imperceptible impairment) as defined in Table 3.3. The DI is a quality indicator like the ODG except for its higher sensitivity towards very low signal qualities.

Table 3.3 ODG Grade scale

Impairment Description	**ITU-R Grade**	**ODG**
Imperceptible	5.0	0.0
Perceptible, but not annoying	4.0	-1.0
Slightly annoying	3.0	-2.0
Annoying	2.0	-3.0
Very annoying	1.0	-4.0

The display of the measurement settings shown on the right side of each diagram is depicted in Figure 3.3. The meaning of the values is as shown in Table 3.4.

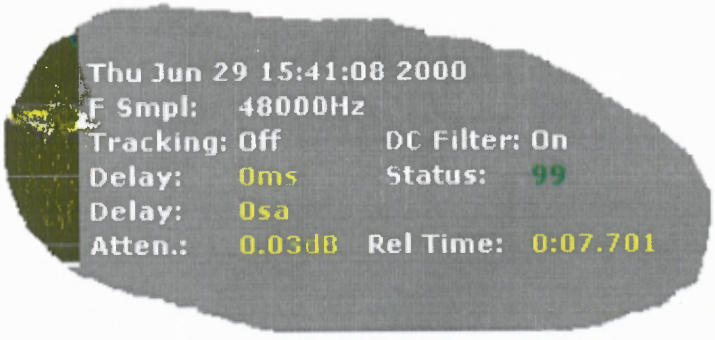


Figure 3.3 General Information related to current display settings.

Table 3.4 Interpretation of the displayed values

Displayed Values	Interpretation
Time	The time when the measurement has been finished.
F Smpl	Sample rate of input signals
Tracking	Status of the delay tracking function (on or off)
DC Filter	Status of the DC filter (on or off)
Delay	Delay in ms (first from top) as well as in samples
Status	Reliability of the automatic delay compensation (0..100%)
Atten	Level difference between reference and test signal (dB)
Rel Time	Current point of time in the measurement

CHAPTER 4

IMPLEMENTATION

This Chapter puts together the design methods used in this thesis. Modifications are made to the architecture depicted in Chapter 2 sensitive processing to take care of structures and properties of audio signals

4.1 Method 1 – DFT based Dead Zone Quantization

The architecture for the first method is shown in Figure 4.1. In this method, zero zone quantization is applied over the DFT transform coefficients of the signal. Also, the presence of voiced and unvoiced bands is taken into consideration. As the human auditory system is more sensitive to the voiced band, introducing distortions into this band is more prominent in the subjective tests. Hence a low threshold is applied in the voiced band and a higher threshold is applied in the unvoiced band.

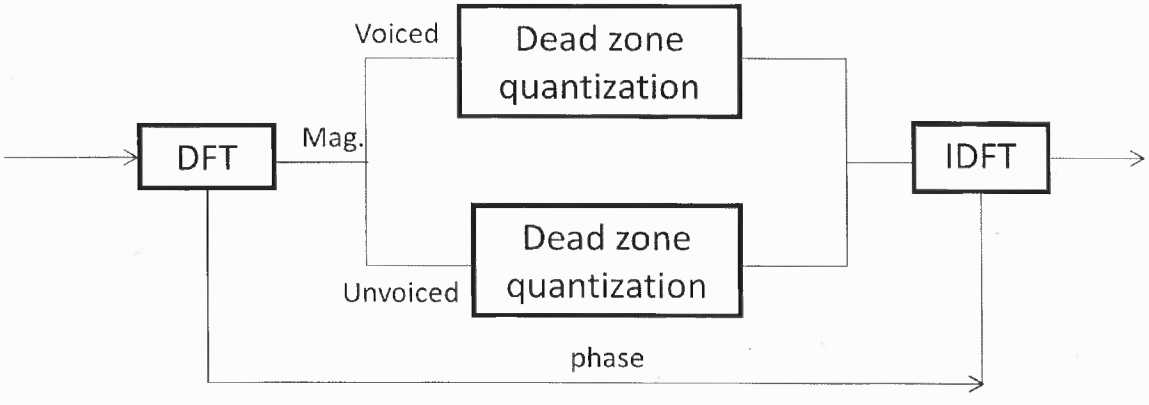


Figure 4.1 Architecture for Method 1.

The threshold that is being applied in method 1 is based on the spread of the coefficients. It is defined as:

$$\lambda = \mu + \alpha \, . \, \sigma \tag{4.3}$$

Where, μ is the mean, σ is the standard deviation and α is a constant. The threshold will increase as α increases.

Since, different threshold values are considered for the voiced and unvoiced bands, two variables determine the quality of output obtained as a result of applying this thresholding: α_1 for the voiced band and α_2 for the unvoiced band. The ODG values and SD were calculated for various thresholds applied to the noisy signal. Tables 4.1 and 4.2 show results for Method 1 using DFT of size 128 and 512 respectively.

Table 4.1 Results for Method 1, DFT size 128

Thresholds		**Results**	
α_1	**α_2**	**ODG**	**SD (dB)**
Input signal		*-3.98*	*8.3887*
-0.2	1.0	-3.93	5.2547
-0.2	3.0	-3.69	1.9821
0	3.0	-3.63	2.0000
0.2	3.0	-3.62	1.9884
0.2	4.0	-3.59	1.8086

Table 4.2 Results for Method 1, DFT size 512

Thresholds		**Results**	
α_1	α_2	**ODG**	**SD (dB)**
Input signal		*-3.98*	*8.3887*
-0.2	3	-3.98	10.9378
0	3	-3.92	4.5086
0.2	3	-3.73	4.3656
0.2	4	-3.59	3.6222
0.2	7	-3.48	3.5204
0.1	7	-3.45	3.5944
0	7	-3.38	3.6942

The ODG value of the noisy signal was measured to be -3.98. Using size 128 DFT, the ODG came up to -3.59 and using 512 sized DFT, the value came up to -3.38. Effectively more noise got removed when using a DFT of greater size. Figure 4.2 shows the waveforms and spectrograms of a noisy signal and best outputs from method 1. Musical notes can be heard in the outputs when processed on the fourier domain. These musical notes appear in the spectrograms as dots in high frequency areas.

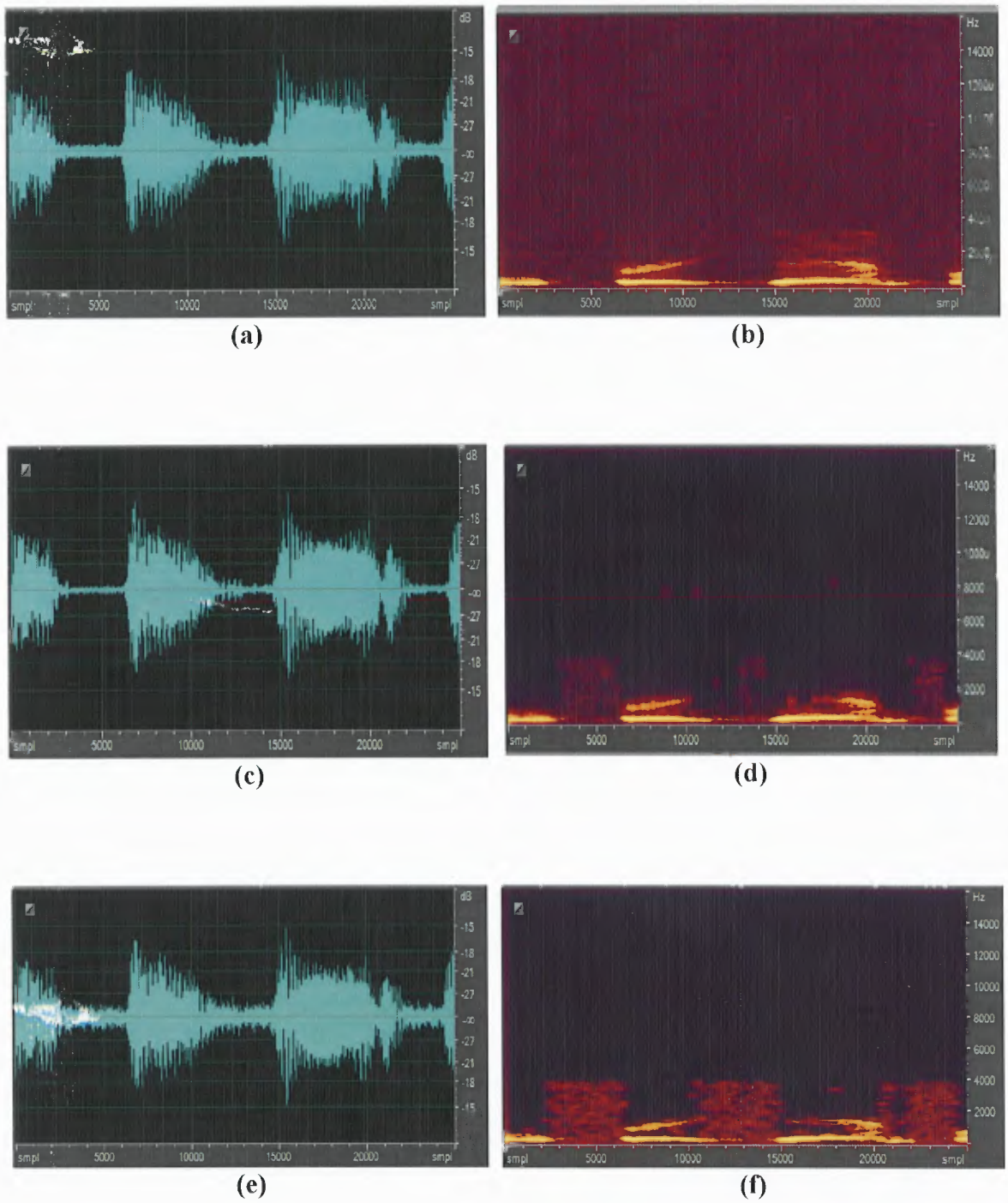


Figure 4.2 (a) Waveform and (b) Spectrogram of noisy signal (c) Waveform and (d) Spectrogram of the best output from method 1 and DFT size 128 (e) Waveform and (f) Spectrogram of the best output from method 1 and DFT size 512.

4.2 Method 2- KLT based Zero Zone Quantization

The input signal is passed through a low pass and high pass filter to separate voiced and unvoiced bands. The KLT of both the bands is taken and dead zone quantization applied separately on both these bands (Figure 4.3).

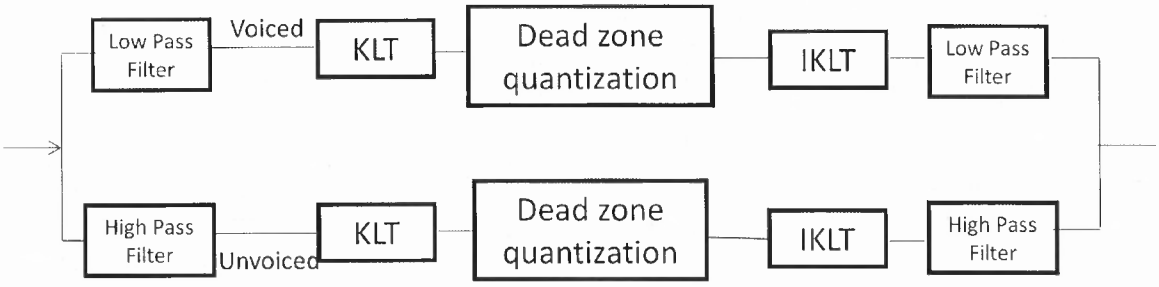


Figure 4.3 Architecture for Method – 2.

In this method, threshold is applied based on the energy content of the transform coefficients. As was discussed in Chapter 2, KLT decorrelates the signal and rearranges the energy content of the input signal. KLT arranges the transform coefficients in the order of decreasing energy of the signal. The threshold decides the amount of energy that will be rejected. A threshold 0.01, for example, means that 1% of the total energy in the signal is getting rejected. Again, α_1 is used for the voiced band and α_2 for the unvoiced band, higher α_2 is applied to get rigorous thresholding in the unvoiced band. For implementation purpose a butterworth filter of 12^{th} order is used for the low pass and the high pass filter.

Tables 4.3 and 4.4 depict the output measures for size 128 and 512 sized KLT, respectively. A higher threshold is being applied to the unvoiced band to get a perceivable output.

Table 4.3 Results for Method 2, KLT size 128

Thresholds		**Results**	
$Energyth_L$	**$Energyth_H$**	**ODG**	**SD(dB)**
Input signal		*-3.98*	*8.3887*
0.0005	0.01	-3.97	8.6564
0.0005	0.8	-3.97	5.8137
0.0005	0.95	-3.98	4.3327
0.0005	0.98	-3.58	4.3587
0.005	0.98	-3.53	4.3542
0.01	0.98	-3.51	4.3387
0.04	0.98	-3.57	4.3286

Table 4.4 Results for Method 2, KLT size 512

Thresholds		**Results**	
$Energyth_L$	**$Energyth_H$**	**ODG**	**SD(dB)**
Input Signal		*-3.98*	*8.3887*
0.0005	0.95	-3.98	4.7107
0.0005	0.98	-3.98	4.4870
0.01	0.98	-3.68	4.4183
0.04	0.98	-3.61	4.3949
0.05	0.98	-3.61	4.4588
0.04	0.99	-3.60	4.3674
0.04	1.0	-3.58	4.3013

The ODG value of the noisy signal was measured to be -3.98. Using size 128 KLT, the ODG came up to -3.51 and using 512 sized KLT, the value came up to -3.58. Theoretically, an improvement is expected as the KLT size increases but as the size of

KLT increases, the numerical complexity of calculating the eigen vectors increases which accounts for numerical inefficiencies.

An improvement in terms of ODG value is visible for 128 sized KLT compared to 128 sized DFT (method 1). The improvement is clearly visible in the waveform and spectrogram plots of Figure 4.4.

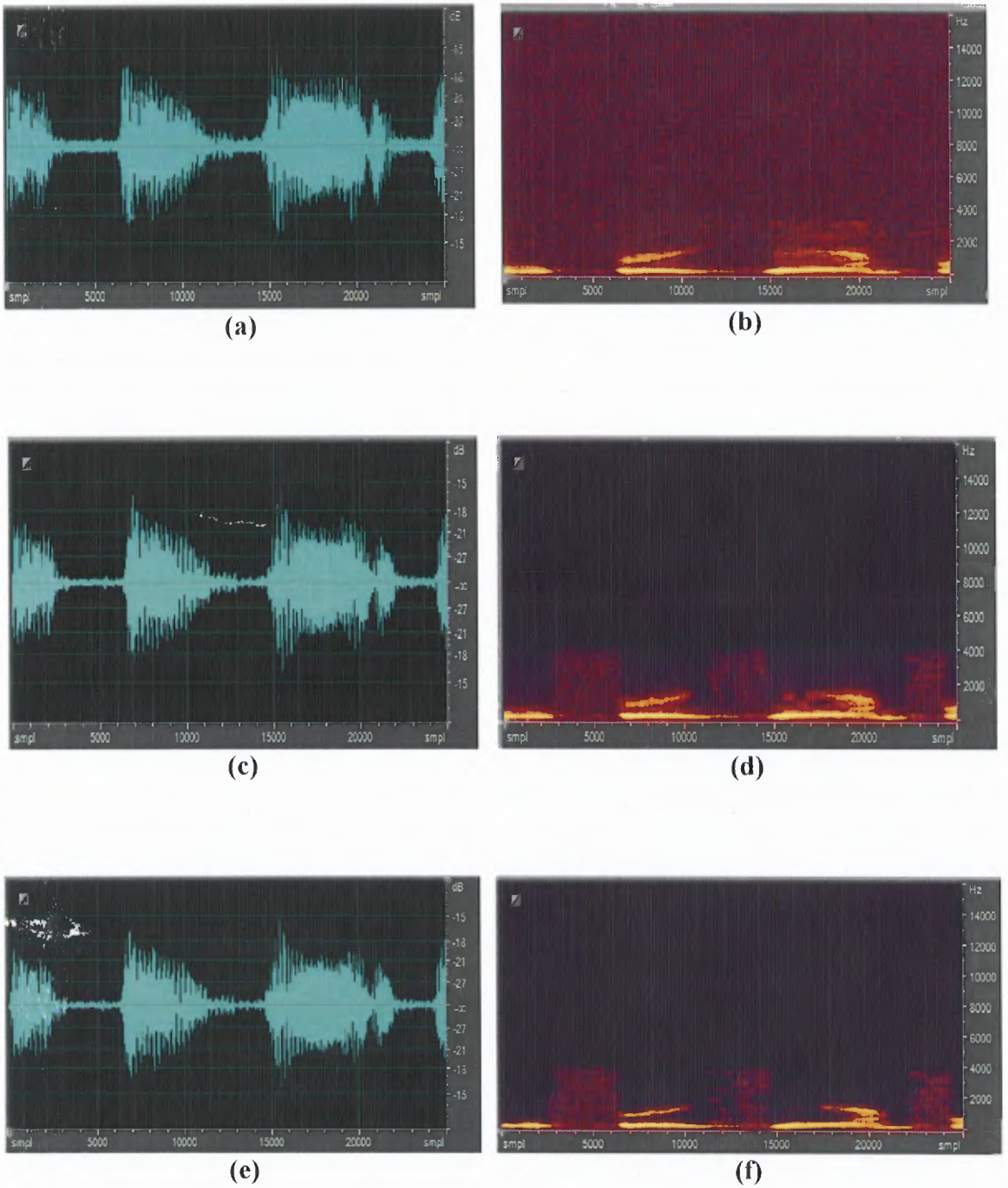


Figure 4.4 (a) Waveform and (b) Spectrogram of noisy signal (c) Waveform and (d) Spectrogram of the best output from method 1 and KLT size 128 (e) Waveform and (f) Spectrogram of the best output from method 1 and KLT size 512.

4.3 Method-3 DFT applied on Dyadic trees

The zero zone quantizer is applied over the DFT transform coefficients from each frequency subband coming out of the dyadic tree (Figure 4.5). As the depth of the tree increases, the threshold gets localized to lower frequency more. Since, the auditory system is more sensitive to low frequency; it becomes quite intuitive that a small threshold be applied.

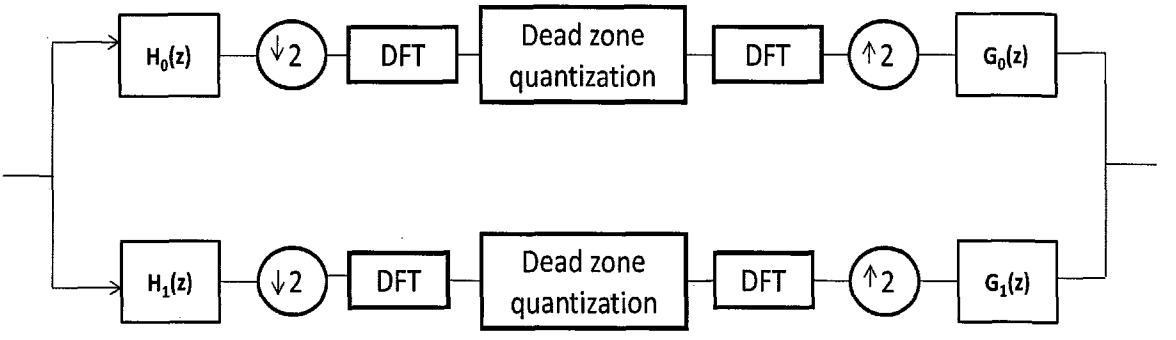


Figure 4.5 Architecture for Method – 3.

Table 4.5 Results for Method 3, Depth – 1 Dyadic Tree

Threshold		128 - DFT		512 - DFT	
α_1	α_2	ODG	SD (dB)	ODG	SD(dB)
4	-0.3	-3.90	5.7138	-2.65	5.6881
4	-0.2	-3.98	5.6349	-2.97	5.6258
4	-0.1	-3.96	5.5793	-3.92	5.4998
4	0	-3.72	5.4986	-3.65	5.3769
4	0.1	-3.87	5.5204	-3.51	5.2649
4	0.2	-3.94	5.5520	-3.53	5.2072
5	0	-3.69	5.5701	-3.55	5.0620

Tables 4.5 and 4.6 summarize the results of applying such a threshold over a depth one and depth 3 dyadic tree, respectively. A higher depth of the dyadic tree indicates better resolution in lower frequency and noise reduction by thresholding becomes more effective.

Table 4.6 Results for Method 3, Depth – 3 Dyadic Tree

Threshold				**128 - DFT**		**512 - DFT**	
α_1	**α_2**	**α_3**	**α_4**	**ODG**	**SD(dB)**	**ODG**	**SD(dB)**
4	4	4	noTh	-3.55	4.1038	-3.32	4.1206
4	4	4	-0.2	-3.55	3.9817	-3.55	3.9434
4	4	3	-0.2	-3.53	3.9615	-3.52	3.9566
4	4	3	-0.1	-3.55	4.0035	-3.55	4.0946
4	4	2.5	-0.2	-3.56	3.9691	-3.50	4.0720

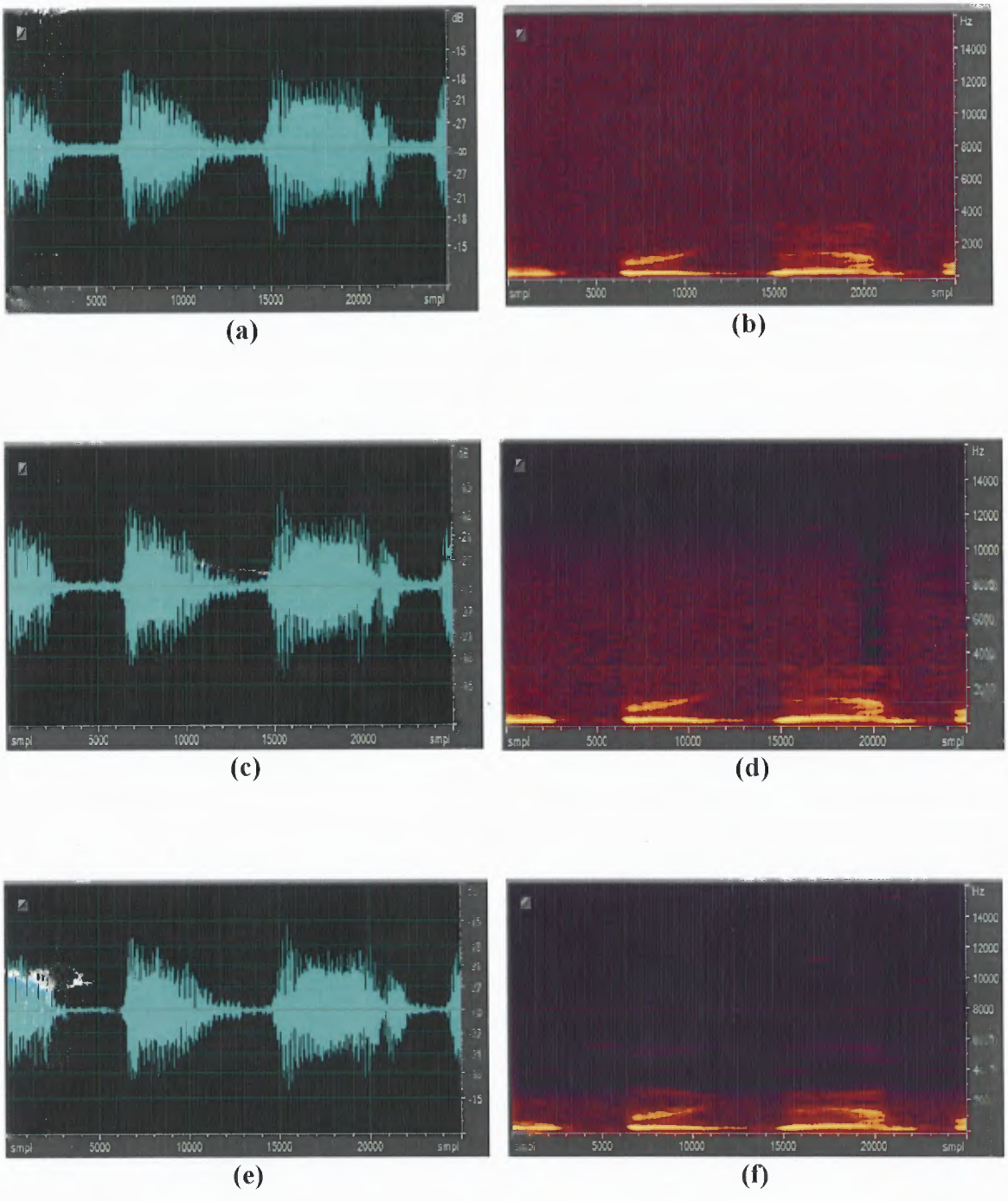


Figure 4.6 (a) Waveform and (b) Spectrogram of noisy signal (c) Waveform and (d)Spectrogram of the best output from method 3 and depth 1 dyadic tree (e) Waveform and (f) Spectrogram of the best output from method 3 and depth 3 dyadic tree.

4.4 Method-4 KLT after Subband decomposition

The zero zone quantizer is applied over the KLT coefficients from each frequency subband coming out of the dyadic tree (Figure 4.7). As the depth of the tree increases, the threshold gets localized to lower frequency. Figure 4.7 shows the architecture for this method.

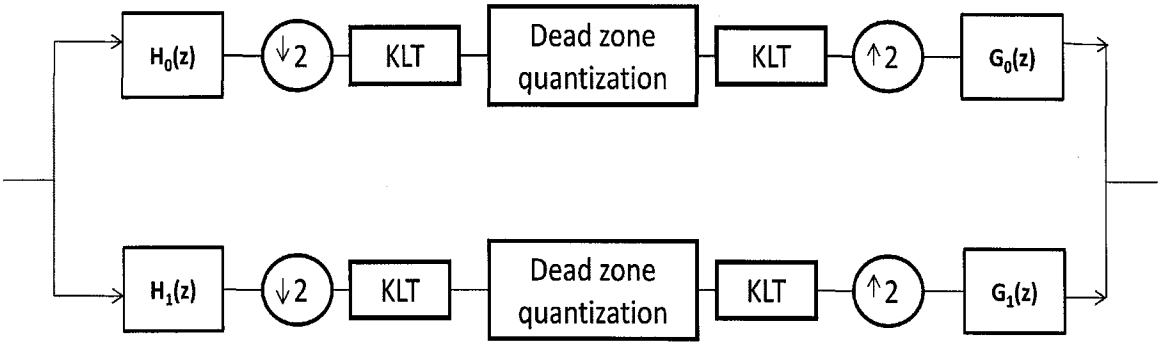


Figure 4.7 Architecture for Method – 4.

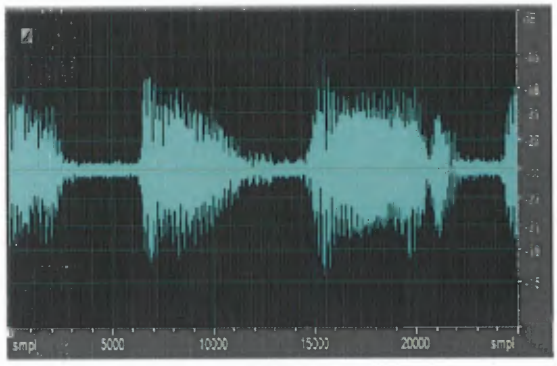
Tables 4.7 and 4.8 put together the results of method 4 for depth one and a depth 3 dyadic tree, respectively. Figure 4.8 shows the waveforms and spectrograms of a noisy signal and best outputs for depth 1 and depth 3 dyadic trees.

Table 4.7 Results for Method 4, Depth – 1 Dyadic Tree

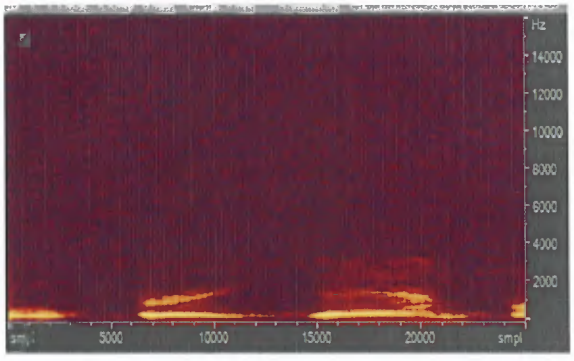
Thresholds		**128 - KLT**		**512 - KLT**	
Energyth$_1$	**Energyth$_2$**	**ODG**	**SD (dB)**	**ODG**	**SD (dB)**
0.99	0.02	-3.62	5.8963	-3.59	5.9488
0.98	0.01	-3.96	6.0264	-3.84	6.0962
0.95	0.01	-3.96	6.0264	-3.88	6.2393
0.99	0.01	-3.84	5.8952	-3.77	6.0325

Table 4.8 Results for Method 4, Depth – 3 Dyadic Tree

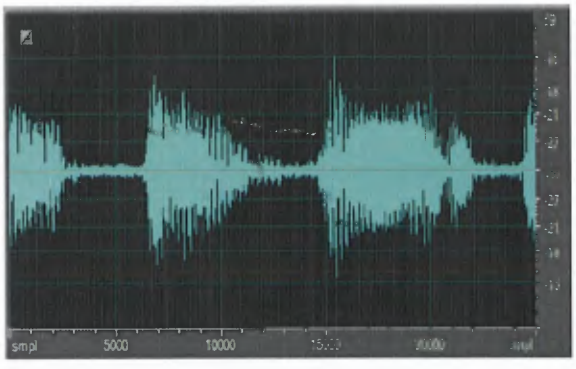
Thresholds				128 - KLT		512 - KLT	
Th_1	Th_2	Th_3	Th_4	ODG	SD (dB)	ODG	SD (dB)
0.99	0.99	0.98	noTh	-3.32	4.1111	-3.54	4.1019
0.99	0.99	0.99	0.01	-3.32	4.1592	-3.52	4.1138
0.99	0.99	0.98	0.01	-3.33	4.1322	-3.54	4.1124
0.99	0.99	0.99	noTh	-3.32	4.1509	-3.52	4.0980



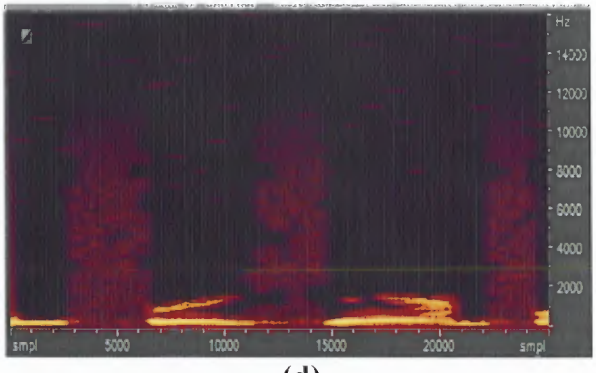
(a)



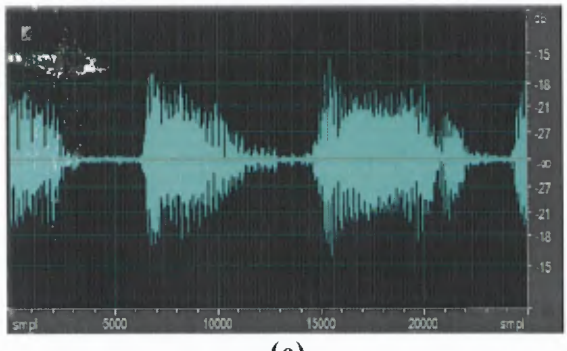
(b)



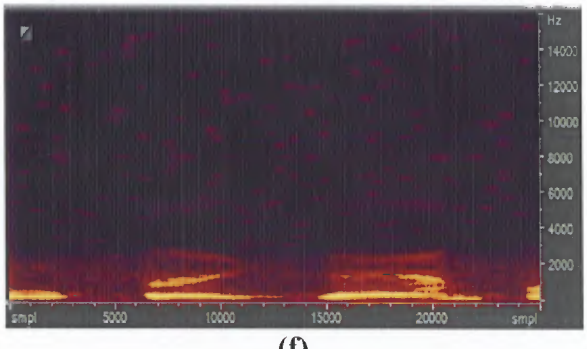
(c)



(d)



(e)



(f)

Figure 4.8 (a) Waveform and (b) Spectrogram of noisy signal (c) Waveform and (d)Spectrogram of the best output from method 4 and depth 1 dyadic tree (e) Waveform and (f) Spectrogram of the best output from method 4 and depth 3 dyadic tree.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusions

This thesis has presented several noise removal methods depending on the threshold applied over the orthogonal transforms. Here, the four proposed methods have proved that Zero Zone Quantization is a simple, yet robust way of achieving noise removal and compression.

Four different concepts were proposed and experiments were carried out. The first two were based on orthogonal transforms, DFT and KLT. For the next two, orthogonal transforms were taken after decomposing the signal into different frequency bands by using Subband decomposition. Different thresholds were applied and experiments were carried out to obtain audio signals with reduced noise content. SD and PEAQ scores were measured based on which the signal was said to have lost noise compared to the input signal.

In context of orthogonal transforms, better noise removal was observed with bigger size of the transform. This validates that higher transform size, decorrelates the signal in a better manner.

Another concept that was suggested and validated experimentally was that applying the transform after subband decomposition gave better results for both Fourier transform and Karhunen Loeve. Subband Decomposition critically samples the input and splits it in frequency scale.

The methods results using KLT gave better results compared to those obtained from processing over the DFT coefficients. However, KLT computations are way too

complex. There are no fast computation algorithms present for KLT as there are for DFT. KLT is completely signal dependent which prevents its application in real life applications. DFT on the other hand are easy to handle. The fast fourier transforms further provide faster implementation of this transforms.

5.2 Future Work

Even though noise was removed from the signal, better results can be achieved by using VADs for the estimation of noise. VAD algorithms are used for feature extraction of the signal based on the previous frames. They can be used to detect the behavior of the noise present in the signal and thus noise can be removed more intelligently.

The Dyadic tree structures used in this work used Daubechies filters to decompose the signal. Use of these filters introduces aliasing due to non zero transition bandwidth and stopband gain. However the overlap in the stopband is highly desired the absence of which would introduce severe attenuation around $\pi/2$ frequency range. Special filters can be designed that do not introduce as much aliasing.

The use of these additional features in noise removal algorithm will make the processing more reliable and accurate.

REFERENCES

[1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 27, pp. 113-120, 1979

[2] *Sparse Representation for Signal Processing and Coding,* Retrieved on August 28, 2009 from http://www.see.ed.ac.uk/~tblumens/Sparse/Sparse.html

[3] A. Gersho and R. Gray, *Vector quantization and signal compression*: Kluwer, 1993.

[4] Wikipedia.org. *Quantization(signal processing),* Retrieved on October 16, 2009 from http://en.wikipedia.org/wiki/Quantization_(signal_processing).

[5] W. Kleijn and K. Paliwal, *Speech coding and synthesis*: Elsevier Science Inc. New York, NY, USA, 1995.

[6] A.N. Akansu and R.A. Haddad, Multiresolution Signal Decomposition: Transforms, Subbands, wavelets, 2rd Edition, Academic Press, 2001.

[7] *Principal Component Analysis and Karhunen- Loeve Transform,* Retrieved on March 23, 2009 from http://fourier.eng.hmc.edu/e161/lectures/klt/

[8] *Optimal Decorrelation and the KLT,* Retrieved on March 23, 2009 form http://www-ee.uta.edu/dip/Courses/EE5356/KLT.doc.

[9] J. Burl, "Estimating the basis functions of the Karhunen-Loeve transform," *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 37, pp. 99-105, 1989.

[10] R. Crochiere and L. Rabiner, *Multirate digital signal processing*: Prentice-Hall Englewood Cliffs, NJ, 1983.

[11] Zou, X., P. Jancovic, et al. (2008). "Speech Signal Enhancement Based on MAP Algorithm in the ICA Space." IEEE Transactions on Signal Processing **56**: 1812-1820.

[12] *User manual for the OPERA™* , Retrieved on September 12, 2009 from http://www.opticomm.de/

[13] Wikipedia.org. *PEAQ,* Retrieved on September 16, 2009 from http://en.wikipedia.org/wiki/PEAQ.

[14] M. Ramkumar and A. Akansu, "A Performance Study of DCT and Subband Image Codecs With Zero-zone Quantizers," 1998, pp. 118-121.

[15] P. Vaidyanathan, *Multirate systems and filter banks*: Dorling Kindersley, licensees of Pearson Education in South Asia, 1993.

[16] S. Chang, *et al.*, "Image denoising via lossy compression and wavelet thresholding," 1997.

[17] M. Zibulevsky and B. Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural computation,* vol. 13, pp. 863-882, 2001.